COOLEY LLP
BOBBY GHAJAR (198719)
(bghajar@cooley.com)
COLETTE GHAZARIAN (322235)
(cghazarian@cooley.com)
1333 2nd Street, Suite 400
Santa Monica, California 90401
Telephone: (310) 883-6400

MARK WEINSTEIN (193043)
(mweinstein@cooley.com)
KATHLEEN HARTNETT (314267)
(khartnett@cooley.com)
JUDD LAUTER (290945)
(jlauter@cooley.com)
ELIZABETH L. STAMESHKIN (260865)
(lstameshkin@cooley.com)
3175 Hanover Street
Palo Alto, CA 94304-1130
Telephone: (650) 843-5000

CLEARY GOTTLIEB STEEN & HAMILTON LLP
ANGELA L. DUNNING (212047)
(adunning@cgsh.com)
1841 Page Mill Road, Suite 250
Palo Alto, CA 94304
Telephone: (650) 815-4131

PAUL, WEISS, RIFKIN, WHARTON & GARRISON LLP
KANNON K. SHANMUGAM (*pro hac vice*)
(kshanmugam@paulweiss.com)
2001 K Street, NW
Washington, DC 20006
Telephone: (202) 223-7300

Counsel for Defendant Meta Platforms, Inc.

UNITED STATES DISTRICT COURT

NORTHERN DISTRICT OF CALIFORNIA

SAN FRANCISCO DIVISION

RICHARD KADREY, *et al.*,
 	
    Individual and Representative Plaintiffs,
 	
    v.
 	
META PLATFORMS, INC., a Delaware corporation;
 	
                                    Defendant.
	Case No. 3:23-cv-03417-VC-TSH
	
	DECLARATION OF MELANIE KAMBADUR IN SUPPORT OF META'S REPLY IN SUPPORT OF ITS MOTION FOR PARTIAL SUMMARY JUDGMENT

1 I, Melanie Kambadur, declare:

2   1.   I am over the age of 18 and am competent to make this declaration. I am a Research

3 Engineering Manager in the Generative AI ("Gen AI") organization of Meta Platforms, Inc.

4 ("Meta"). I have been employed by Meta since March 2021. I have personal knowledge of the

5 facts contained in this declaration in support of Defendant Meta Platform Inc.'s Motion for Partial

6 Summary Judgment. I declare that the following is true to the best of my knowledge, information,

7 and belief, and that if called upon to testify, I could and would testify to the following.

8 <center>**Professional Background**</center>

9   2.   In 2010, I received a Bachelor's degree in Computer Science from Indiana

10 University Bloomington, and in 2016, I received a Ph.D. from Columbia University in Computer

11 Science. Prior to joining Meta, I worked as a software engineer and engineering manager at Oscar

12 Health from 2016 through 2020, and was a co-founder and CTO at Podnods Incorporated, a podcast

13 discovery startup, from 2020-2021.

14   3.   My work at Meta began as part of an organization previously called Facebook AI

15 Research ("FAIR"), which has now changed its name to Fundamental AI Research. I moved into

16 the Gen AI organization in approximately March of 2023. My employment at Meta has included

17 research and development of Meta's machine learning and large language model (LLM)

18 technologies, including the Llama LLMs, but also Meta LLMs that predate Llama. For Llama 2

19 onward, I have been one of the managers overseeing model development. My responsibilities have

20 included supporting researchers and engineers on the Llama team, including research relating to

21 post-training and pre-training. I am familiar with Meta's research and development of its LLMs,

22 and with the ways in which training datasets are used internally within Meta to develop and train

23 its LLMs.

24 <center>**Use of Datasets**</center>

25   4.   Meta has used a number of text datasets to train its Llama large language models

26 (LLMs), including portions of datasets that include books. Text training datasets undergo pre-

27 processing to make them more suitable for training models which can include, for example, de-

28 duplicating the data to reduce the number of times the model is exposed to the same data during

training. Meta's use of text datasets goes beyond using them to train models that are *directly* released or made available to the public.

5.   For example, Meta also uses text training datasets for internal research purposes in connection with the development of its LLMs. Meta has used these text datasets to perform controlled experiments (often referred to as "ablations") for evaluation of the data contained in them, in order to determine the probable impact on performance of the model if they were used in training. In one common type of ablation, Meta uses an internal LLM that was created for experiments (which is often a smaller version of a pre-release Llama model), and, using industry benchmarks, measures the model's performance before and after a particular dataset was added to its training mix. This allows Meta to assess the value (if any) the additional dataset will provide for LLM training. In late 2022, for example, researchers within the FAIR organization used data from LibGen to perform experiments to determine the potential impact of the data on model performance based on several industry benchmarks.

6.   The internal use of datasets for research purposes is essential to the development of LLMs because it enables researchers to better understand and identify how particular types of data impact LLM performance. This research, for example, not only helps identify the types of data that improve model performance, but also allows researchers to spot types of data that diminish performance or result in biased responses.

7.   The internal use of text training datasets for research and development purposes is essential to Meta's development of its LLMs, including its Llama LLMs. I am unaware of any use by Meta of text training datasets comprised of books that was not related to AI research, development or training.

//

//

//

//

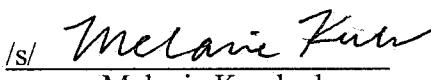**DECLARATION OF MELANIE KAMBADUR**
CASE NO. 3:23-CV-03417-VC-TSH

1   I declare under penalty of perjury that the foregoing is true and correct. Executed on this

2 ⟋⟋ day of April, at New York , New York .

/s/ *Melanie Kun*

Melanie Kambadur

3

4

5

6

7

8

9

10

11

12

13

14

15

16

17

18

19

20

21

22

23

24

25

26

27

28